

# On Sets of Large Fourier Transform Under Changes in Domain

Joel Laity and Barak Shani

Department of Mathematics, University of Auckland, New Zealand

## Abstract

A function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  can be represented as a linear combination  $f(x) = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \chi_{\alpha,n}(x)$  where  $\hat{f}$  is the (discrete) Fourier transform of  $f$ . Clearly, the basis  $\{\chi_{\alpha,n}(x) := \exp(2\pi i \alpha x / n)\}$  depends on the value  $n$ .

We show that if  $f$  has “large” Fourier coefficients, then the function  $\tilde{f} : \mathbb{Z}_m \rightarrow \mathbb{C}$ , given by

$$\tilde{f}(x) = \begin{cases} f(x) & \text{when } 0 \leq x < \min(n, m), \\ 0 & \text{otherwise,} \end{cases}$$

also has “large” coefficients. Moreover, they are all contained in a “small” interval around  $\lfloor \frac{m}{n} \alpha \rfloor$  for each  $\alpha \in \mathbb{Z}_n$  such that  $\hat{f}(\alpha)$  is large. One can use this result to recover the large Fourier coefficients of a function  $f$  by redefining it on a convenient domain. One can also use this result to reprove a result by Morillo and Ràfols: *single-bit* functions, defined over any domain, have a small set of large coefficients.

## 1. INTRODUCTION

In recent years different areas of mathematics, such as additive number theory [2], combinatorial number theory [5] and cryptography [4], have seen results that take advantage of the structure arising from large values of the Fourier transform. Most notably, the quantum algorithm for period finding given by Shor [6] is an application that exploits this structure.

In this paper we are concerned with what happens to the Fourier coefficients of a function when we extend or restrict its domain. Our convention is to set the function to be zero on the new added values.

**Definition 1.** Let  $m, n \in \mathbb{N}$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$ . Define  $\tilde{f} : \mathbb{Z}_m \rightarrow \mathbb{C}$  by

$$\tilde{f}(x) = \begin{cases} f(x) & \text{when } 0 \leq x < \min(n, m), \\ 0 & \text{otherwise.} \end{cases}$$

In this paper we will establish a relationship between the large Fourier coefficients of  $f$  and  $\tilde{f}$ . Our results are motivated by the following:

1. When  $m > n$ , our results have some relevance to approximations of functions.

There is a broad area of research, mostly in the computer science and engineering communities,

that studies algorithms to recover the set of points for which the Fourier transform of a given function is large (see [1] for a recent survey). Such algorithms have been extensively studied and optimized for domains of size a power of 2.

Therefore, our motivation is to show that if the large Fourier coefficients of  $\tilde{f}$  can be recovered for  $m = 2^k$ , one can then determine the set of large Fourier coefficients of  $f$ .

2. When  $m < n$ , our results apply to functions that can be naturally defined over any domain and the families of functions they give rise to  $\{f_n : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$ .

One example of such a family is to take a periodic function  $f : \mathbb{Z} \rightarrow \mathbb{C}$  and restrict its domain by defining  $f_n(x) = f(x)$  for all  $x \in \mathbb{Z}_n$ . Another example is the family of most-significant-bit (MSB) functions: for  $2^k < n \leq 2^{k+1}$ ,  $\text{MSB}(x) = 0$  if  $x < 2^k$  and  $\text{MSB}(x) = 1$  otherwise. It turns out that the analysis of the Fourier transform for these functions is much easier on some domains than others.

Thus, our motivation is to show that by analyzing the large Fourier coefficients of only a small proportion of functions in the family, one can analyze the large Fourier coefficients of all the functions in the family.

## 1.1 Definitions and Results

We first give some definitions as well as some basic properties of the Fourier transform.

We use  $\lfloor x \rfloor$  to denote the closest integer to  $x \in \mathbb{R}$  and we use  $\|x\|$  for the distance between  $x$  and its closest integer, i.e.  $\|x\| = |\lfloor x \rfloor - x|$ . Let  $k \in \mathbb{N}$ , for all  $x \in \mathbb{R}$  define  $|x|_k = \min\{|x - kz| \mid z \in \mathbb{Z}\}$ , so  $|x|_k$  is the distance from  $x$  to the nearest integer multiple of  $k$ . Equivalently we can define  $|x|_k = k\|x/k\|$ .

The group  $\mathbb{Z}_n$  is the set  $\{0, 1, \dots, n-1\}$  under the group operation of addition modulo  $n$ . The set  $L^2(\mathbb{Z}_n) = \{f : \mathbb{Z}_n \rightarrow \mathbb{C}\}$  of all functions from  $\mathbb{Z}_n$  to  $\mathbb{C}$  forms a vector space under the usual addition and scalar multiplication of functions. This space has the inner product  $\langle f, g \rangle = \frac{1}{n} \sum_{x \in \mathbb{Z}_n} f(x) \overline{g(x)}$ , where  $\bar{z}$  denotes the complex conjugate of  $z \in \mathbb{C}$ , and norm  $\|f\|_2 = \sqrt{\langle f, f \rangle}$ .

For each  $\alpha \in \mathbb{Z}_n$  the additive character  $\chi_{\alpha,n}$  is given by  $\chi_{\alpha,n}(x) = \omega_n^{\alpha x}$ , where  $\omega_n = \exp(2\pi i/n)$  is the  $n$ -th root of unity. When it is clear from context what the domain of a character is, we will omit the second index and use  $\chi_\alpha$  to denote  $\chi_{\alpha,n}$ . The set of all characters  $\{\chi_\alpha\}_{\alpha \in \mathbb{Z}_n}$  forms an orthonormal basis for  $L^2(\mathbb{Z}_n)$ ; thus any function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  can be written as

$$f(x) = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \chi_\alpha(x), \quad (1)$$

where  $\hat{f}(\alpha) \in \mathbb{C}$  is the  $\alpha$ -th *Fourier coefficient*. Equation (1) implicitly defines a function  $\hat{f} : \mathbb{Z}_n \rightarrow \mathbb{C}$ ,

called the Fourier transform of  $f$ . It can be explicitly calculated using the formula

$$\widehat{f}(\alpha) = \langle f, \chi_\alpha \rangle = \frac{1}{n} \sum_{x \in \mathbb{Z}_n} f(x) \overline{\chi_\alpha(x)}.$$

One of the most fundamental identities in Fourier analysis, and one used throughout this paper, is Parseval's identity:

$$\frac{1}{n} \sum_{x \in \mathbb{Z}_n} |f(x)|^2 = \|f\|_2^2 = \sum_{\alpha \in \mathbb{Z}_n} |\widehat{f}(\alpha)|^2.$$

Given a function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  and a set  $\Gamma \subseteq \mathbb{Z}_n$  we use the notation  $f|_\Gamma$  to denote the projection of  $f$  onto the subspace  $\text{span}\{\chi_\alpha \mid \alpha \in \Gamma\} \subseteq L^2(\mathbb{Z}_n)$ ; that is  $f|_\Gamma$  is a function from  $\mathbb{Z}_n$  to  $\mathbb{C}$  and  $f|_\Gamma(x) := \sum_{\alpha \in \Gamma} \widehat{f}(\alpha) \chi_\alpha(x)$ . Note that this definition of  $f|_\Gamma$  is different from the usual meaning of the notation.

A function is  $k$ -sparse if  $\widehat{f}(\alpha)$  is nonzero for at most  $k$  values of  $\alpha \in \mathbb{Z}_n$ . In this paper we are concerned with two types of functions. The first is of functions with Fourier coefficients whose magnitude is large *relative to the function's norm*. The second type is of functions that can be approximated, up to any error term, in the form  $f|_\Gamma$  where  $\Gamma$  has small cardinality. We call these functions concentrated.

**Definition 2.** Let  $\{n_k\}_{k \in \mathbb{N}}$  be a sequence of positive integers. Let  $\epsilon > 0$ . A family of functions  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is  $\epsilon$ -concentrated if there exists a polynomial  $P \in \mathbb{R}[x]$  such that for all  $k \in \mathbb{N}$  there exists a set  $\Gamma_k \subseteq \mathbb{Z}_{n_k}$  where  $|\Gamma_k| \leq P(\log(n_k))$  and  $\|f_k - f_k|_{\Gamma_k}\|_2^2 < \epsilon$ .

**Definition 3.** Let  $\{n_k\}_{k \in \mathbb{N}}$  be a sequence of positive integers. A family of functions  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is concentrated if there exists a bivariate polynomial  $P \in \mathbb{R}[x, y]$  such that for all  $k \in \mathbb{N}$  and for all  $\epsilon > 0$  there exists a set  $\Gamma_k \subseteq \mathbb{Z}_{n_k}$  where  $|\Gamma_k| \leq P(\log(n_k), 1/\epsilon)$  and  $\|f_k - f_k|_{\Gamma_k}\|_2^2 < \epsilon$ .

Roughly speaking, Definition 3 says that a family of functions is concentrated if it can be approximated up to any error term by polynomially many characters (with respect to the domain size and the approximation error).

Most applications are concerned with a single function that implicitly defines the entire family for all possible domains (that is,  $n_k = k$  or  $n_k = c + k$  for some constant  $c$ ). In this case we abuse notation and say that the function, instead of the family, is concentrated. Here are some examples.

### Example 1.

- A single character defines a concentrated family. More precisely if we fix  $\alpha \in \mathbb{N}$  the family of functions  $\{\chi_\alpha : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n > \alpha}$  is concentrated. For every integer  $n > \alpha$  we can choose  $\Gamma_n = \{\alpha\}$ , then  $\|\chi_\alpha - \chi_\alpha|_{\Gamma_n}\|_2^2 = 0$  and hence the definition is satisfied with the constant polynomial  $P = 1$ .

- *Single-bit functions on domains of size powers of 2 are concentrated families. More precisely, let  $k \in \mathbb{N}$  and  $0 \leq i < k$ . Define  $\text{bit}_i : \mathbb{Z}_{2^k} \rightarrow \{-1, 1\}$  by  $\text{bit}_i(x) = (-1)^{x_i}$  where  $x = \sum_{j=0}^{k-1} x_j 2^j$  and  $x_j \in \{0, 1\}$ . Then, as we show in Lemma 6.2 below, the family  $\mathcal{F}_i = \{\text{bit}_i : \mathbb{Z}_{2^k} \rightarrow \{-1, 1\}\}_{k \geq i}$  is concentrated for every  $i \in \mathbb{N}$ .*

Moreover, as a corollary of our result in Section 6.2, we show that single-bit functions are concentrated over every domain.

- *The family  $\{f_n : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$  where  $\widehat{f_n}(\alpha) = 1$  and  $\widehat{f_n}(\alpha') = 1/\sqrt{n}$  for all  $\alpha' \neq \alpha$  is  $\epsilon$ -concentrated for  $\epsilon = 1$ . We can choose  $\Gamma_n = \{\alpha\}$  for every  $n \in \mathbb{N}$  then  $\|f_n - f_n|_{\Gamma_n}\|_2^2 = (n-1)(\frac{1}{\sqrt{n}})^2 = 1 - 1/n < \epsilon$ . However this family is not  $\epsilon$ -concentrated for any  $\epsilon < 1$  since even if  $\alpha \in \Gamma_n$  we have  $\|f_n - f_n|_{\Gamma_n}\|_2^2 = \sum_{\alpha' \in \mathbb{Z}_n, \alpha' \notin \Gamma_n} |\widehat{f_n}(\alpha')|^2 = (n - |\Gamma_n|)(\frac{1}{\sqrt{n}})^2 = 1 - |\Gamma_n|/n$ . Hence if  $|\Gamma_n|$  is polynomial in  $\log(n)$  then  $|\Gamma_n|/n < 1 - \epsilon$  for sufficiently large  $n$ , so  $\|f_n - f_n|_{\Gamma_n}\|_2^2 = 1 - |\Gamma_n|/n > \epsilon$ .*

Functions with similar Fourier transform to this example arose for example from ‘noisy’ characters, given by  $f(x) := \omega_n^{\alpha x + e(x)}$  for some suitable random bounded functions  $e$ .

In this work, we study how the large Fourier coefficients of the function  $f$  are related to those of  $\widetilde{f}$ . It is clear that when  $m$  is much smaller than  $n$ , we lose a lot of information about the function  $f$  (as shown in Example 3 below). In order to avoid extreme cases of unbalanced functions<sup>1</sup>, in our main theorems we restrict the domain  $\mathbb{Z}_m$  such that  $m \geq n/2$ . We remark that our results also hold when  $m \geq n/\text{poly}(\log(n))$ .<sup>2</sup>

Our first result treats concentrated families.

**Theorem 1.1.** *Let  $\{n_k\}_{k \in \mathbb{N}}, \{m_k\}_{k \in \mathbb{N}}$  two sequences of positive integers with  $m_k \geq n_k/2$  for every  $k \in \mathbb{N}$ . Let  $Q \in \mathbb{R}[x]$  be a polynomial. Let  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  be a concentrated family of functions such that  $\|f_k\|_2^2 \leq Q(\log(n_k))$  for all  $k \in \mathbb{N}$ . Then  $\{\widetilde{f}_k : \mathbb{Z}_{m_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is a concentrated family of functions.*

Furthermore, we show that there is a simple relationship between the sets  $\Gamma_k$  on which  $f_k$  are concentrated and the sets on which  $\widetilde{f}_k$  are concentrated. This leads to an algorithmic approach to approximate a function  $f$ , which we treat in Section 6.1: by finding an approximation of  $\widetilde{f}$  with a sparse function, one can translate it to approximate  $f$  with a sparse function.

Our second result treats any function  $f$  with large Fourier coefficients.

<sup>1</sup>An example of such extreme case is a function where  $f(0) = n$  and otherwise  $f(x) = 0$ ; then  $\|f\|_2^2 = 1$ , and if one takes  $m = 1$  then  $\|\widetilde{f}\|_2^2 = n$ .

<sup>2</sup>Notice that by taking  $m \gg n$  some of the results hold trivially.

**Theorem 1.2.** Let  $\{n_k\}_{k \in \mathbb{N}}, \{m_k\}_{k \in \mathbb{N}}$  two sequences of positive integers with  $m_k \geq n_k/2$  for every  $k \in \mathbb{N}$ . Let  $t := \sup_{k \in \mathbb{N}} \{n_k/m_k\}$ . Let  $Q \in \mathbb{R}[x]$  be a polynomial. Let  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  be a family of functions such that  $\|f_k\|_2^2 \leq Q(\log(n_k))$  for all  $k \in \mathbb{N}$ . Then if  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is an  $\epsilon$ -concentrated family of functions, the family  $\{\tilde{f}_k : \mathbb{Z}_{m_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is a  $(t\epsilon + \eta)$ -concentrated family of functions for any  $\eta > 0$ .

## 2. PREPARATION

We first show how to express the Fourier transform of  $\tilde{f}$  in terms of the Fourier coefficients of  $f$ .

**Lemma 2.1.** Let  $m, n \in \mathbb{N}$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$ . Let  $\ell := \min(n, m)$ . Let  $g : \mathbb{Z}_m \rightarrow \mathbb{C}$  be defined by  $g(x) = f(x)$  for  $0 \leq x < \ell$  and  $g(x) = 0$  otherwise. Then

$$\hat{g}(\beta) = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x} \quad (2)$$

for all  $\beta \in \mathbb{Z}_m$ .

*Proof.* Recall that  $f(x) = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \chi_\alpha(x)$ . So

$$\begin{aligned} \hat{g}(\beta) &= \frac{1}{m} \sum_{x \in \mathbb{Z}_m} g(x) \overline{\omega_m^{\beta x}} = \frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} f(x) \omega_m^{-\beta x} \\ &= \frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \omega_n^{\alpha x} \omega_m^{-\beta x} = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x}. \end{aligned}$$

□

Lemma 2.1 tells us that the Fourier coefficient  $\hat{g}(\beta)$  is a weighted sum of the Fourier coefficients of  $f$ . The weights are given by the exponential sum  $\frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x}$ , the magnitude of which is inversely proportional to the distance between  $\frac{m}{n}\alpha$  and  $\beta$  (we will give a precise statement and proof in Lemma 3.1). Hence if  $\beta$  is very far from  $\frac{m}{n}\alpha$  then the  $\hat{f}(\alpha)$ -term in Equation (2) will contribute very little to  $\hat{g}(\beta)$ .

**Lemma 2.2.** Let  $m, n \in \mathbb{N}$ . The mapping  $\tilde{\cdot} : L^2(\mathbb{Z}_n) \rightarrow L^2(\mathbb{Z}_m)$  is linear. Moreover,

$$\langle \tilde{f}, \tilde{g} \rangle = \frac{n}{m} \langle f, g \rangle \quad \text{if } n < m, \quad \text{and} \quad \|\tilde{f}\|_2^2 \leq \frac{n}{m} \|f\|_2^2$$

for all  $f, g : \mathbb{Z}_n \rightarrow \mathbb{C}$ , with equality if  $n < m$ .

*Proof.* The linearity follows directly from the definition of  $\tilde{\cdot}$ . From the definition of the inner product  $\langle \tilde{f}, \tilde{g} \rangle = \frac{1}{m} \sum_{x=0}^{m-1} \tilde{f}(x) \overline{\tilde{g}(x)} = \frac{1}{m} \sum_{x=0}^{n-1} f(x) \overline{g(x)} = \frac{n}{m} \langle f, g \rangle$ , when  $n < m$ . In addition, we always have  $\langle \tilde{f}, \tilde{f} \rangle = \frac{1}{m} \sum_{x=0}^{m-1} \tilde{f}(x) \overline{\tilde{f}(x)} \leq \frac{1}{m} \sum_{x=0}^{n-1} f(x) \overline{f(x)} = \frac{n}{m} \langle f, f \rangle$ . The last property now follows since  $\|\tilde{f}\|_2^2 = \langle \tilde{f}, \tilde{f} \rangle \leq \frac{n}{m} \langle f, f \rangle = \frac{n}{m} \|f\|_2^2$ . □

**Remark 1.** Another version of Lemma 2.1 is

$$\widehat{g}(\beta) = \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle.$$

We can also give an alternative proof that uses the properties of  $\widetilde{\cdot}$  instead of explicit calculations. From the linearity of  $\widetilde{\cdot}$  we get  $g = \widetilde{f} = \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \widetilde{\chi}_{\alpha,n}$ . Then

$$\widehat{g}(\beta) = \langle g, \chi_{\beta,m} \rangle = \left\langle \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \right\rangle = \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle.$$

Thus  $\widehat{g}(\beta)$  is weighted sum of the Fourier coefficients of  $f$  where the weights are  $\langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle$ .

In order to bound the exponential sum in Equation (2) we will need the following lemma.

**Lemma 2.3.** Let  $N \in \mathbb{N}$ . Let  $x \in \mathbb{R} \setminus \mathbb{Z}$ . Let  $e(x) := \exp(2\pi i x)$ . Then

$$\sum_{k=0}^N e(xk) = e(xN/2) \frac{\sin(\pi x(N+1))}{\sin(\pi x)}.$$

*Proof.* From the formula for a geometric series

$$\sum_{k=0}^N e(xk) = \frac{e(x(N+1)) - 1}{e(x) - 1}.$$

It is possible to pull convenient factors out of the numerator and denominator so that what is left is purely imaginary

$$\frac{e(x(N+1)) - 1}{e(x) - 1} = \frac{e(x(N+1)/2)}{e(x/2)} \frac{e(x(N+1)/2) - e(-x(N+1)/2)}{e(x/2) - e(-x/2)}.$$

We substitute  $y = x(N+1)/2$  into the formula  $\sin(2\pi y) = (e(y) - e(-y))/(2i)$  and get

$$\sum_{k=0}^N e(xk) = e(xN/2) \frac{\sin(\pi x(N+1))}{\sin(\pi x)},$$

as required. □

**Lemma 2.4.** Let  $k \in \mathbb{N}$ . Let  $x \in \mathbb{R}$  such that  $x$  is not an integer multiple of  $k$ . Then

$$\frac{1}{|\sin(\frac{\pi}{k}x)|} \leq \frac{k}{2|x|_k}.$$

*Proof.* Applying the elementary bound<sup>3</sup>  $|\sin(\pi x)| \geq 2\|x\|$  for all  $x \in \mathbb{R}$  we get  $1/|\sin(\pi x/k)| \leq 1/(2\|x/k\|) = 1/(2|x|_k/k)$  as required. □

---

<sup>3</sup>The easiest way to see why this is true is by drawing the graphs of both functions.

### 3. FUNCTIONS WITH A SINGLE COEFFICIENT

We start our analysis by considering functions  $f$  with only one non-zero coefficient, i.e.  $f(x) = \widehat{f}(\alpha)\chi_\alpha(x)$ . In Lemma 3.1 we give an upper bound on the size of each Fourier coefficient of  $\widetilde{f}$  and a lower bound on the size of the  $\lfloor \frac{m}{n}\alpha \rfloor$ -th Fourier coefficient of  $\widetilde{f}$ .

**Lemma 3.1.** *Let  $m, n \in \mathbb{N}$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  be a 1-sparse function, i.e.  $f(x) = \widehat{f}(\alpha)\chi_\alpha(x)$  for some  $\alpha \in \mathbb{Z}_n$ . Let  $\ell := \min(n, m)$ . Let  $g : \mathbb{Z}_m \rightarrow \mathbb{C}$  be defined by  $g(x) = f(x)$  for  $0 \leq x < \ell$  and  $g(x) = 0$  otherwise. If  $\beta = \frac{m}{n}\alpha$  then  $|\widehat{g}(\beta)| = \frac{\ell}{m}|\widehat{f}(\alpha)|$ . For all  $\beta \in \mathbb{Z}_m$  with  $\beta \neq \frac{m}{n}\alpha$  we have*

$$|\widehat{g}(\beta)| \leq |\widehat{f}(\alpha)| \min \left( \frac{\ell}{m}, \frac{1}{2 \lfloor \frac{m}{n}\alpha - \beta \rfloor_m} \right). \quad (3)$$

Moreover if  $\beta = \lfloor \frac{m}{n}\alpha \rfloor$  then

$$|\widehat{g}(\beta)| \geq |\widehat{f}(\alpha)| \max \left( \frac{2\ell}{\pi m}, \frac{\ell}{m} \sqrt{1 - \frac{\pi^2 \ell^2}{3m^2} \left\lfloor \frac{m}{n}\alpha - \beta \right\rfloor_m^2} \right). \quad (4)$$

*Proof.* From Lemma 2.1

$$|\widehat{g}(\beta)| = |\widehat{f}(\alpha)| \frac{1}{m} \left| \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x} \right| = |\widehat{f}(\alpha)| \frac{1}{m} \left| \sum_{x=0}^{\ell-1} \exp \left( \frac{2\pi i x}{m} \left( \frac{m}{n}\alpha - \beta \right) \right) \right| \leq |\widehat{f}(\alpha)| \frac{\ell}{m}.$$

Since  $\alpha \in \mathbb{Z}_n$  and  $\beta \in \mathbb{Z}_m$  the expression  $(\frac{m}{n}\alpha - \beta)/m$  is an integer if and only if  $\beta = \frac{m}{n}\alpha$ . In this case we get that  $|\widehat{g}(\beta)| = |\widehat{f}(\alpha)| \frac{\ell}{m}$ . Otherwise, from Lemma 2.3

$$|\widehat{g}(\beta)| = |\widehat{f}(\alpha)| \frac{1}{m} \left| \frac{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \ell \right)}{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \right)} \right|. \quad (5)$$

We bound the right-hand side of this equation using

$$\frac{1}{m} \left| \frac{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \ell \right)}{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \right)} \right| \leq \frac{1}{m} \frac{1}{\left| \sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \right) \right|} \leq \frac{1}{2 \lfloor \frac{m}{n}\alpha - \beta \rfloor_m}, \quad (6)$$

where the last inequality follows from substituting  $x = \frac{m}{n}\alpha - \beta$  into Lemma 2.4. This proves (3).

For the second part let  $r = \lfloor \frac{m}{n}\alpha - \beta \rfloor_m$  where  $\beta = \lfloor \frac{m}{n}\alpha \rfloor$ . Then  $0 < r \leq 1/2$ . We now substitute our new variable,  $r$ , into the right-hand side of Equation (5) and bound the resulting expression with the elementary inequalities  $\frac{2}{\pi}|x| \leq |\sin x|$  for all  $x \in [-\pi/2, \pi/2]$  and  $|\sin x| \leq |x|$  for all  $x \in \mathbb{R}$ . We get

$$\frac{1}{m} \left| \frac{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \ell \right)}{\sin \left( \frac{\pi}{m} \left( \frac{m}{n}\alpha - \beta \right) \right)} \right| = \frac{1}{m} \left| \frac{\sin \left( \frac{\pi \ell}{m} r \right)}{\sin \left( \frac{\pi}{m} r \right)} \right| \geq \frac{2\ell}{\pi m}.$$

To get a lower bound that depends on  $r$  we first observe that, since  $\sin^2(x) = (1 - \cos 2x)/2$  holds for all  $x \in \mathbb{R}$ , the Maclaurin series of  $\sin^2(x)$  is  $x^2 - x^4/3 + \text{high order terms}$ . Hence  $x^2 - x^4/3 \leq \sin^2(x)$  for all  $x \in \mathbb{R}$ . Combining this with the elementary bound  $\sin^2(x) \leq x^2$  we get

$$\frac{1}{m^2} \left| \frac{\sin\left(\frac{\pi \ell}{m} r\right)}{\sin\left(\frac{\pi}{m} r\right)} \right|^2 \geq \frac{1}{m^2} \frac{\left(\frac{\pi \ell}{m} r\right)^2 - \frac{\left(\frac{\pi \ell}{m} r\right)^4}{3}}{\left(\frac{\pi}{m} r\right)^2} = \frac{\ell^2}{m^2} \left(1 - \frac{\pi^2 \ell^2}{3m^2} r^2\right).$$

So  $|\widehat{g}(\beta)| \geq |\widehat{f}(\alpha)| \frac{\ell}{m} \sqrt{1 - \frac{\pi^2 \ell^2}{3m^2} r^2}$  as required. This proves Equation (4) and completes the proof.  $\square$

**Remark 2.** One natural question is whether we can get any nontrivial lower bound on  $|\widehat{g}(\beta)|$  where  $\beta$  is not the closest integer to  $\frac{m}{n}\alpha$ . First notice that in the proof we rely on  $r = |\frac{m}{n}\alpha - \beta|_m$  being at most  $1/2$  in order to use the bound  $|\sin x| \geq \frac{2}{\pi}|x|$ . Since this holds for at most two  $\beta$ , we can not use the same idea to lower bound other coefficients of  $g$ .

We first notice that if  $m$  and  $n$  are coprime and  $\beta \neq \frac{m}{n}\alpha$  then  $\|r\| \geq \frac{1}{n}$ . Using  $2\|x\| \leq \sin(\pi x) \leq \pi\|x\|$  for all  $x \in \mathbb{R}$  we get

$$\frac{1}{m} \left| \frac{\sin\left(\frac{\pi}{m} r \ell\right)}{\sin\left(\frac{\pi}{m} r\right)} \right| \geq \frac{2}{\pi} \frac{\|\ell r/m\|}{m\|r/m\|} = \frac{2}{\pi} \frac{\|\ell r/m\|}{|r|_m} \geq \begin{cases} 2/(\pi m|r|_m), & \text{if } n < m, \\ 2/(\pi n|r|_m), & \text{if } m < n. \end{cases}$$

It turns out we can not do much better than this bound, as

$$\frac{1}{m} \left| \frac{\sin\left(\frac{\pi}{m} r \ell\right)}{\sin\left(\frac{\pi}{m} r\right)} \right| \leq \frac{\pi}{2} \frac{\|\ell r/m\|}{m\|r/m\|} = \frac{\pi}{2} \frac{\|\ell r/m\|}{|r|_m}.$$

Recall  $r = |\frac{m}{n}\alpha - \beta|_m$ , and if we take  $m = n + 1$ , then we get  $r = |\alpha - \beta + \frac{1}{n}\alpha|_m$ . If we take  $\beta = \lfloor \frac{m}{n}\alpha \rfloor - 1 = \alpha - 1$  and  $\alpha = 2$ , then  $|\widehat{f}(\alpha)|/(\pi m) \leq |\widehat{g}(\beta)| \leq |\widehat{f}(\alpha)|\pi/m$ .

#### 4. PROOFS OF THE MAIN THEOREMS

In this section we prove a sequence of propositions from which Theorem 1.1 and Theorem 1.2 will follow. To illustrate the ideas we first consider a single character  $\chi_\alpha$ , and show that  $\widetilde{\chi}_\alpha$  has small number of large coefficients. Secondly, we show the same result for a  $k$ -sparse function. Lastly, we consider any function with  $k$  large coefficients and give the same result.

**Proposition 1.** Let  $m, n \in \mathbb{N}$ . Let  $\epsilon > 0$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  be a character, i.e.  $f(x) = \chi_\alpha(x)$  for some  $\alpha \in \mathbb{Z}_n$ . Let  $\ell := \min(n, m)$ . Let  $g : \mathbb{Z}_m \rightarrow \mathbb{C}$  be defined by  $g(x) = f(x)$  for  $0 \leq x < \ell$  and  $g(x) = 0$  otherwise. Let  $\Gamma'$  be the interval in  $\mathbb{Z}_m$  defined by  $\Gamma' = \{\beta \in \mathbb{Z}_m \mid |\frac{m}{n}\alpha - \beta|_m \leq r + 1\}$  where  $r = 1/(2\epsilon)$ . Then

$$\|g - g|_{\Gamma'}\|_2^2 \leq \epsilon.$$



*Proof.* We begin by using the bound from Lemma 3.1

$$\|g - g|_{\Gamma'}\|_2^2 = \sum_{\beta \notin \Gamma'} |\widehat{g}(\beta)|^2 \leq \sum_{\beta \notin \Gamma'} \frac{1}{4|\frac{m}{n}\alpha - \beta|_m^2}.$$

The set  $\Gamma'$  consists of all elements of  $\mathbb{Z}_m$  whose distance is at most  $r + 1$  from  $\frac{m}{n}\alpha$ . Thus the set  $\mathbb{Z}_m \setminus \Gamma'$  can be written as  $L \cup R$  where  $L$  consists of all elements “to the left” of  $\frac{m}{n}\alpha$ , i.e.  $L = \{\lceil \frac{m}{n}\alpha \rceil - x \mid x = r + 1, r + 2, \dots, \lceil m/2 \rceil\} \subseteq \mathbb{Z}_m$  and  $R$  is all elements “to the right” of  $\frac{m}{n}\alpha$ , i.e.  $R = \{\lfloor \frac{m}{n}\alpha \rfloor + x \mid x = r + 1, r + 2, \dots, \lceil m/2 \rceil\} \subseteq \mathbb{Z}_m$ . Then our sum becomes

$$\|g - g|_{\Gamma'}\|_2^2 \leq \sum_{\beta \in L \cup R} \frac{1}{4|\frac{m}{n}\alpha - \beta|_m^2} \leq 2 \sum_{x=r+1}^{\lceil m/2 \rceil} \frac{1}{4x^2} \leq \frac{1}{2r} = \epsilon.$$

where in the last inequality we use the fact that  $\sum_{x=r+1}^{\infty} \frac{1}{x^2} \leq \frac{1}{r}$ .  $\square$

**Proposition 2.** Let  $m, n \in \mathbb{N}$ . Let  $\epsilon > 0$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  be a  $|\Gamma|$ -sparse function, i.e.  $f(x) = \sum_{\alpha \in \Gamma} \widehat{f}(\alpha) \chi_\alpha(x)$ . Let  $\ell := \min(n, m)$ . Let  $g : \mathbb{Z}_m \rightarrow \mathbb{C}$  be defined by  $g(x) = f(x)$  for  $0 \leq x < \ell$  and  $g(x) = 0$  otherwise. Let  $\Gamma' = \bigcup_{\alpha \in \Gamma} \{\beta \in \mathbb{Z}_m \mid |\frac{m}{n}\alpha - \beta|_m \leq r + 1\}$  for  $r = |\Gamma| \|f\|_2^2 / (2\epsilon)$ . Then

$$\|g - g|_{\Gamma'}\|_2^2 \leq \epsilon.$$

*Proof.* We can bound  $|\widehat{g}(\beta)|$  with the triangle inequality

$$|\widehat{g}(\beta)| = |\langle g, \chi_{\beta, m} \rangle| = \left| \left\langle \sum_{\alpha \in \Gamma} \widehat{f}(\alpha) \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \right\rangle \right| \leq \sum_{\alpha \in \Gamma} |\widehat{f}(\alpha)| |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle|.$$

Then

$$\|g - g|_{\Gamma'}\|_2^2 = \sum_{\beta \notin \Gamma'} |\widehat{g}(\beta)|^2 \leq \sum_{\beta \notin \Gamma'} \left| \sum_{\alpha \in \Gamma} |\widehat{f}(\alpha)| |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle| \right|^2.$$

The Cauchy–Schwarz inequality gives

$$\begin{aligned} \|g - g|_{\Gamma'}\|_2^2 &\leq \sum_{\beta \notin \Gamma'} \left( \sum_{\alpha \in \Gamma} |\widehat{f}(\alpha)|^2 \sum_{\alpha \in \Gamma} |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle|^2 \right) \\ &= \sum_{\beta \notin \Gamma'} \left( \|f\|_2^2 \sum_{\alpha \in \Gamma} |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle|^2 \right) \\ &= \|f\|_2^2 \sum_{\alpha \in \Gamma} \sum_{\beta \notin \Gamma'} |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle|^2. \end{aligned}$$

For the last part of the proof we bound  $\sum_{\beta \notin \Gamma'} |\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle|^2$  using Proposition 1 with  $\epsilon$  replaced by  $\frac{\epsilon}{|\Gamma| \|f\|_2^2}$  (as  $\langle \widetilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle = \widehat{\widetilde{\chi}}_{\alpha, n}(\beta)$ ). This gives

$$\|g - g|_{\Gamma'}\|_2^2 \leq \|f\|_2^2 \sum_{\alpha \in \Gamma} \frac{\epsilon}{|\Gamma| \|f\|_2^2} = \epsilon.$$

$\square$

**Proposition 3.** Let  $m, n \in \mathbb{N}$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$ . Suppose  $\Gamma \subseteq \mathbb{Z}_n$  is such that

$$\|f - f|_{\Gamma}\|_2^2 \leq \epsilon$$

for some  $\epsilon > 0$ . Let  $\ell := \min(n, m)$ . Let  $g : \mathbb{Z}_m \rightarrow \mathbb{C}$  be defined by  $g(x) = f(x)$  for  $0 \leq x < \ell$  and  $g(x) = 0$  otherwise. Let  $\epsilon' > 0$ . Let  $\Gamma' = \bigcup_{\alpha \in \Gamma} \{\beta \in \mathbb{Z}_m \mid |\frac{m}{n}\alpha - \beta|_m \leq r + 1\}$  for  $r = |\Gamma| \|f\|_2^2 / (2\epsilon')$ . Then

$$\|g - g|_{\Gamma'}\|_2^2 \leq t\epsilon + \epsilon' + 2\sqrt{t\epsilon\epsilon'},$$

where  $t = n/m$ .

*Proof.* We write  $f = f|_{\Gamma} + f|_{\mathbb{Z}_n \setminus \Gamma}$ . From linearity of the  $\widetilde{\cdot}$  operator

$$g = \widetilde{f} = \widetilde{f}|_{\Gamma} + \widetilde{f}|_{\mathbb{Z}_n \setminus \Gamma}.$$

There is a possibility for some confusion with the notation. By  $\widetilde{f}|_{\Gamma}$  we mean  $(\widetilde{f|_{\Gamma}})$ , not  $\widetilde{f}$  projected on  $\Gamma$  (the latter is not defined as  $\Gamma \subseteq \mathbb{Z}_n$ ). By linearity of the Fourier Transform

$$\widehat{g}(\beta) = \widehat{\widetilde{f}}|_{\Gamma}(\beta) + \widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)$$

for all  $\beta \in \mathbb{Z}_m$ . Hence

$$\begin{aligned} \|g - g|_{\Gamma'}\|_2^2 &= \sum_{\beta \notin \Gamma'} |\widehat{g}(\beta)|^2 = \sum_{\beta \notin \Gamma'} \left| \widehat{\widetilde{f}}|_{\Gamma}(\beta) + \widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta) \right|^2 \\ &\leq \sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\Gamma}(\beta)|^2 + 2 \sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\Gamma}(\beta)| |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)| + \sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)|^2. \end{aligned}$$

The rest of the proof consists of bounding the three terms on the right-hand side of the above equation. Since  $f|_{\Gamma} : \mathbb{Z}_n \rightarrow \mathbb{C}$  is  $|\Gamma|$ -sparse we can apply Proposition 2 with  $\epsilon'$ , which says

$$\sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\Gamma}(\beta)|^2 = \|\widetilde{f}|_{\Gamma} - (\widetilde{f}|_{\Gamma})|_{\Gamma'}\|_2^2 \leq \epsilon'.$$

By hypothesis  $\|f|_{\mathbb{Z}_n \setminus \Gamma}\|_2^2 = \|f - f|_{\Gamma}\|_2^2 \leq \epsilon$ . Hence

$$\sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)|^2 \leq \sum_{\beta \in \mathbb{Z}_n} |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)|^2 = \|\widetilde{f}|_{\mathbb{Z}_n \setminus \Gamma}\|_2^2 \leq \frac{n}{m} \|f|_{\mathbb{Z}_n \setminus \Gamma}\|_2^2 \leq t\epsilon.$$

Finally by the Cauchy–Schwarz inequality

$$2 \sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\Gamma}(\beta)| |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)| \leq 2 \sqrt{\sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\Gamma}(\beta)|^2 \sum_{\beta \notin \Gamma'} |\widehat{\widetilde{f}}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)|^2} \leq 2\sqrt{\epsilon' t \epsilon}.$$

Putting it all together we get

$$\begin{aligned} \sum_{\beta \notin \Gamma'} |\widehat{g}(\beta)|^2 &\leq \sum_{\beta \notin \Gamma'} |\widehat{f}|_{\Gamma}(\beta)|^2 + 2 \sum_{\beta \notin \Gamma'} |\widehat{f}|_{\Gamma}(\beta)|\widehat{f}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)| + \sum_{\beta \notin \Gamma'} |\widehat{f}|_{\mathbb{Z}_n \setminus \Gamma}(\beta)|^2 \\ &\leq \epsilon' + 2\sqrt{\epsilon' t \epsilon} + t\epsilon. \end{aligned}$$

□

**Proof of Theorem 1.1.** Let  $\epsilon > 0$ . By hypothesis the family  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is concentrated. Hence there exists a bivariate polynomial  $P \in \mathbb{R}[x, y]$  and sets  $\Gamma_k \subseteq \mathbb{Z}_{n_k}$  such that  $|\Gamma_k| \leq P(\log(n_k), \frac{1}{\epsilon/6})$  and  $\|f_k - f_k|_{\Gamma_k}\|_2^2 \leq \epsilon/6$  for all  $k \in \mathbb{N}$ .

By Proposition 3 with  $\epsilon' = \epsilon/6$  there exist sets  $\Gamma'_k \subseteq \mathbb{Z}_{m_k}$  such that  $\|\widetilde{f}_k - \widetilde{f}_k|_{\Gamma'_k}\|_2^2 \leq t_k \epsilon/6 + \epsilon/6 + 2\sqrt{t_k}(\epsilon/6)$ , for all  $k \in \mathbb{N}$ , where  $t_k := n_k/m_k$ . By hypothesis  $n_k/m_k \leq 2$  so

$$\|\widetilde{f}_k - \widetilde{f}_k|_{\Gamma'_k}\|_2^2 \leq 2\epsilon/6 + \epsilon/6 + 2\sqrt{2}(\epsilon/6) < \epsilon$$

for all  $k \in \mathbb{N}$ . All that remains to show is that the sets  $\Gamma'_k$  have polynomial size. Since Proposition 3 gives an explicit construction we know that each  $\Gamma'_k$  is a union of  $|\Gamma_k|$  intervals of width at most  $2(r+1)+1 = 2r+3$  where  $r = |\Gamma_k| \|f_k\|_2^2 / (2\epsilon')$  hence

$$|\Gamma'_k| \leq |\Gamma_k|(2r+3) = |\Gamma_k| \left( 2 \frac{|\Gamma_k| \|f_k\|_2^2}{2\epsilon'} + 3 \right) = |\Gamma_k| \left( \frac{6|\Gamma_k| \|f_k\|_2^2}{\epsilon} + 3 \right)$$

By hypothesis there exists a polynomial  $Q \in \mathbb{R}[x]$  such that  $\|f_k\|_2^2 \leq Q(\log(n_k))$  for all  $k \in \mathbb{N}$ . When we upper bound  $|\Gamma_k|$  with  $P(\log(n_k), 6/\epsilon)$  and upper bound  $\|f_k\|_2^2$  with  $Q(\log(n_k))$  the displayed inequality becomes

$$|\Gamma'_k| \leq P(\log(n_k), 6/\epsilon) (6\epsilon^{-1} P(\log(n_k), 6/\epsilon) Q(\log(n_k)) + 3).$$

By hypothesis  $n_k/2 \leq m_k$ , and so  $\log(n_k) \leq \log(m_k) + 1$ . Therefore, the expression is polynomial in  $1/\epsilon$  and  $\log(m_k)$  as required. So the definition of concentration is satisfied with the polynomial  $P'(x, y) = P(x+1, 6y)(6yP(x+1, 6y)Q(x+1) + 3)$ . □

**Proof of Theorem 1.2.** The proof of this theorem is similar to that of Theorem 1.1. By hypothesis the family  $\{f_k : \mathbb{Z}_{n_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is  $\epsilon$ -concentrated. Hence there exists a polynomial  $P \in \mathbb{R}[x]$  and there exist sets  $\Gamma_k \subseteq \mathbb{Z}_{n_k}$  such that for all  $k \in \mathbb{N}$  we have  $|\Gamma_k| \leq P(\log(n_k))$  and  $\|f_k - f_k|_{\Gamma_k}\|_2^2 \leq \epsilon$ .

From Proposition 3 there exist sets  $\Gamma'_k \subseteq \mathbb{Z}_{m_k}$  such that  $\|\widetilde{f}_k - \widetilde{f}_k|_{\Gamma'_k}\|_2^2 \leq t\epsilon + \epsilon' + 2\sqrt{t\epsilon\epsilon'}$  for all  $k \in \mathbb{N}$  (recall that  $t := \sup_{k \in \mathbb{N}} \{n_k/m_k\}$ ). If we choose  $\epsilon' = \min(\eta/2, \eta^2/(16t\epsilon))$  then we get

$$\|\widetilde{f}_k - \widetilde{f}_k|_{\Gamma'_k}\|_2^2 \leq t\epsilon + \eta/2 + 2\sqrt{t\epsilon \frac{\eta^2}{16t\epsilon}} = t\epsilon + \eta/2 + \eta/2 = t\epsilon + \eta.$$

Moreover, Proposition 3 gives us a bound for  $|\Gamma'_k|$ . Recall that  $\Gamma'_k$  is a union of  $|\Gamma_k|$  intervals of at most width  $2r + 3$  where  $r = |\Gamma_k| \|f_k\|_2^2 / (2\epsilon')$  hence

$$|\Gamma'_k| \leq |\Gamma_k|(2r + 3) = |\Gamma_k| \left( 2 \frac{|\Gamma_k| \|f_k\|_2^2}{2\epsilon'} + 3 \right).$$

By hypothesis there exists a polynomial  $Q \in \mathbb{R}[x]$  such that  $\|f_k\|_2^2 \leq Q(\log(n_k))$  for all  $k \in \mathbb{N}$ . When we upper bound  $|\Gamma_k|$  with  $P(\log(n_k))$  and upper bound  $\|f_k\|_2^2$  with  $Q(\log(n_k))$  the displayed inequality becomes

$$|\Gamma'_k| \leq P(\log(n_k)) (P(\log(n_k))Q(\log(n_k))/\epsilon' + 3).$$

By hypothesis  $n_k/2 \leq m_k$ , and so  $\log(n_k) \leq \log(m_k) + 1$ . Since  $\epsilon'$  does not depend on  $k$  this expression is polynomial in  $\log(m_k)$  as required.<sup>4</sup> So the family  $\{\tilde{f}_k : \mathbb{Z}_{m_k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  is  $(t\epsilon + \eta)$ -concentrated with respect to the polynomial  $P'(x) = P(x+1)(P(x+1)Q(x+1)/\epsilon' + 3)$ .  $\square$

## 5. COEFFICIENT CANCELLATION

In this section we give some examples of functions where  $|\hat{f}(\alpha)|$  is large but  $|\hat{f}(\lfloor \frac{m}{n}\alpha \rfloor)|$  is not. We then show that  $|\hat{f}(\alpha)|$  being large implies  $|\hat{f}(\lfloor \frac{m}{n}\alpha \rfloor)|$  is large if  $f$  satisfies certain conditions.

Given some function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$ , which can be well approximated by  $f|_\Gamma$ , Proposition 3 shows us how to construct a set  $\Gamma' \subseteq \mathbb{Z}_m$  such that  $\tilde{f} : \mathbb{Z}_m \rightarrow \mathbb{C}$  can be well approximated by  $\tilde{f}|_{\Gamma'}$ .

However, the theorem does not guarantee that  $|\hat{f}(\beta)|$  is large for every  $\beta \in \Gamma'$ . So, if we choose  $\Gamma'$  as in the theorem, this will be sufficient to get a good approximation but some of the members of  $\Gamma'$  may not be necessary for that approximation.

In the case where the function  $f$  is a single character we proved a stronger result: if  $f = \chi_\alpha$  then  $\tilde{f} = \tilde{\chi}_\alpha$  and  $\hat{f}(\lfloor \frac{m}{n}\alpha \rfloor)$  is greater than  $2\ell/(\pi m)$  where  $\ell = \min(n, m)$ . Thus if we want to closely approximate  $\tilde{f}$  it is not only sufficient but necessary for  $\Gamma'$  to contain  $\lfloor \frac{m}{n}\alpha \rfloor$ . Since  $\tilde{f} = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \tilde{\chi}_\alpha$  is just a sum of  $\tilde{\chi}_\alpha$  one might expect this result can be generalised to arbitrary functions. However there is no easy way to conclude which coefficients of  $\tilde{f}$  are large, or even how many of them there are. Observe that

$$\hat{\tilde{f}}(\beta) = \sum_{\alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \langle \tilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle = \hat{f}(\alpha') \langle \tilde{\chi}_{\alpha', n}, \chi_{\beta, m} \rangle + \sum_{\alpha' \neq \alpha \in \mathbb{Z}_n} \hat{f}(\alpha) \langle \tilde{\chi}_{\alpha, n}, \chi_{\beta, m} \rangle.$$

So even if  $\hat{f}(\alpha') \langle \tilde{\chi}_{\alpha', n}, \chi_{\beta, m} \rangle$  is large, e.g. when  $\alpha' = \lfloor \frac{m}{n}\alpha \rfloor$ , it may be cancelled out by the other term.

**Example 2.** Let  $n$  be an odd number and consider the function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  given by

$$f(x) := \begin{cases} 1 & \text{when } x \text{ is even,} \\ -1 & \text{when } x \text{ is odd.} \end{cases}$$

---

<sup>4</sup>However, since  $\epsilon'$  depends on  $\epsilon, \eta$  one can redefine a global polynomial  $P'(x, y)$  in those terms to show the affect of  $\eta$  on the polynomial. This is not needed for our proof, as the only important variable is the domain size.

This function has several large coefficients centred at  $n/2$  (follows from Lemma 2.3; see also Eq. (7) in Lemma 6.2 below). Let  $m = n + 1$  and let  $d = m/2 \in \mathbb{N}$ . We show that  $d$ -th Fourier coefficient of  $\tilde{f} : \mathbb{Z}_m \rightarrow \mathbb{C}$  is large and that the magnitude of the rest is  $1/m$ . Notice that for every  $0 \leq x < n$  we have  $\tilde{f}(x) = \chi_{d,m}(x)$ . Therefore

$$\begin{aligned}\widehat{\tilde{f}}(\beta) &= \frac{1}{m} \sum_{x \in \mathbb{Z}_m} \tilde{f}(x) \overline{\chi_\beta(x)} = \frac{1}{m} \sum_{0 \leq x < n} \tilde{f}(x) \overline{\chi_\beta(x)} = \frac{1}{m} \sum_{0 \leq x < n} \chi_{d-\beta,m}(x) \\ &= \frac{1}{m} \sum_{0 \leq x < n} \chi_{d-\beta,m}(x) + \frac{1}{m} \sum_{n \leq x < m} \chi_{d-\beta,m}(x) - \frac{1}{m} \sum_{n \leq x < m} \chi_{d-\beta,m}(x) \\ &= \frac{1}{m} \sum_{0 \leq x < m} \chi_{d-\beta,m}(x) - \frac{1}{m} \sum_{n \leq x < m} \chi_{d-\beta,m}(x).\end{aligned}$$

Therefore, for  $\beta \neq d$ , we get  $|\widehat{\tilde{f}}(\beta)| = |\frac{1}{m} \sum_{0 \leq x < m} \chi_{d-\beta,m}(x) - \frac{1}{m} \chi_{d-\beta,m}(n)| = 0 + \frac{1}{m}$ . Since  $\|\tilde{f}\|_2^2 = \frac{1}{m}(m-1)$  from Parseval's identity we get  $|\widehat{\tilde{f}}(d)| > \frac{1}{m}(m-1) - (m-1)\frac{1}{m^2} = 1 - O(1/m)$ . We therefore see that all large coefficients of  $f$ , except for one, are canceled out in  $\tilde{f}$ . This example also works for  $m = n + 1 + 2k$  for some  $k$  not too large, e.g.  $k \approx \sqrt{n}$ .

**Example 3.** Let  $n$  be even. Let  $m = n/2$ . Define

$$f(x) := \begin{cases} 0 & \text{when } 0 \leq x < m, \\ \chi_{\alpha,n}(x) & \text{when } m \leq x < n. \end{cases}$$

Then  $\widehat{f}(\alpha)$  is large. However  $\tilde{f}$  is the zero function, hence all its Fourier coefficients are zero. In the appendix we refine this example to show that the bound in Theorem 1.2 is tight.

Recall from Lemma 2.1 that  $\widehat{\tilde{f}}(\beta)$  is a weighted sum of the Fourier coefficients of  $f$ ,

$$\widehat{\tilde{f}}(\beta) = \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x}.$$

In addition, the weight of the  $\alpha$ -th Fourier coefficient,  $\frac{1}{m} \sum_{x \in \mathbb{Z}_\ell} \omega_m^{(\frac{m}{n}\alpha - \beta)x}$ , is bounded by  $1/(2|\frac{m}{n}\alpha - \beta|_m)$  (See Equations (5) and (6) in the proof of Lemma 3.1). Therefore if  $f$  has a few large Fourier coefficients and they are all spaced far apart, then the only term in the sum that will contribute significantly to  $\widehat{\tilde{f}}(\beta)$  is the  $\alpha$ -th term where  $|\widehat{f}(\alpha)|$  is large and  $|\frac{m}{n}\alpha - \beta|_m$  is minimal. Proposition 4 formalises this argument.

**Proposition 4.** Let  $m, n \in \mathbb{N}$  with  $n \leq m \leq 2n$ . Let  $L > 0$ . Let  $C > 1$ . Let  $\Gamma \subseteq \mathbb{Z}_n$ . Let  $r = 20C(\log(|\Gamma|/2) + 1)$ . Let  $\tau < \frac{L}{20 \frac{1}{3+2\log(n)}}$ . Let  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$ , and suppose  $\Gamma \subseteq \mathbb{Z}_n$  satisfies  $|\alpha - \alpha'|_n > r$  for all  $\alpha, \alpha' \in \Gamma$  with  $\alpha \neq \alpha'$ , and that

$$L \leq |\widehat{f}(\alpha)| \leq CL$$

for all  $\alpha \in \Gamma$  and

$$|\widehat{f}(\alpha)| \leq \tau$$

for all  $\alpha \notin \Gamma$ . Then

$$|\widehat{f}(\lfloor \frac{m}{n} \alpha \rfloor)| \geq \frac{1}{5}L$$

for all  $\alpha \in \Gamma$ .

*Proof sketch.* Let  $\gamma \in \Gamma$  and  $\beta = \lfloor \frac{m}{n} \gamma \rfloor$ . Then by the reverse triangle inequality

$$\begin{aligned} |\widehat{f}(\beta)| &= \left| \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle \right| \\ &\geq \left| \widehat{f}(\gamma) \langle \widetilde{\chi}_{\gamma,n}, \chi_{\beta,m} \rangle \right| - \left| \sum_{\gamma \neq \alpha \in \Gamma} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle \right| - \left| \sum_{\alpha \notin \Gamma} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle \right|. \end{aligned}$$

By Lemma 3.1 the first term satisfies  $\left| \widehat{f}(\gamma) \langle \widetilde{\chi}_{\gamma,n}, \chi_{\beta,m} \rangle \right| \geq \frac{2}{\pi} \frac{n}{m} L$ . By hypothesis  $n/m \geq 1/2$  so  $\left| \widehat{f}(\gamma) \langle \widetilde{\chi}_{\gamma,n}, \chi_{\beta,m} \rangle \right| \geq \frac{1}{\pi} L$ . Similar to the proof of Proposition 1, we have for the second term

$$\left| \sum_{\gamma \neq \alpha \in \Gamma} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle \right| \leq CL \sum_{\gamma \neq \alpha \in \Gamma} |\langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle| \leq \frac{CL}{2} \sum_{\gamma \neq \alpha \in \Gamma} \frac{1}{|\frac{m}{n} \alpha - \beta|_m}.$$

Since  $\Gamma$  is  $r$ -spread we know that  $\sum_{\gamma \neq \alpha \in \Gamma} \frac{1}{|\frac{m}{n} \alpha - \beta|_m} \leq 2(\frac{1}{r} + \frac{1}{2r} + \frac{1}{3r} \cdots)$  so

$$\left| \sum_{\gamma \neq \alpha \in \Gamma} \widehat{f}(\alpha) \langle \widetilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle \right| \leq CL \sum_{x=1}^{|\Gamma|/2} \frac{1}{xr} \leq \frac{CL}{r} (\log(|\Gamma|/2) + 1) = \frac{1}{20}L.$$

where we use the fact that  $\sum_{x=1}^N \frac{1}{x} < \log(N) + 1$  in the second to last inequality. Similarly, we can bound the third term by  $\frac{1}{20}L$  using  $|\widehat{f}(\alpha)| \leq \tau$  for all  $\alpha \notin \Gamma$ . Then  $|\widehat{f}(\beta)| \geq \frac{1}{\pi}L - \frac{1}{20}L - \frac{1}{20}L \geq \frac{1}{5}L$ .  $\square$

A fully detailed proof can be found in the master's thesis of the first-named author [3].

## 6. APPLICATIONS

We give some applications of our results. In Section 6.1 we show that our results can be turned into an algorithm to recover the large coefficients of a function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  using the large coefficients of the corresponding function  $\widetilde{f} : \mathbb{Z}_{2^k} \rightarrow \mathbb{C}$ . Algorithms for finding large coefficients on the latter domain have been highly optimized and are generally simpler than the algorithms that work over any domain.

In Section 6.2 we give a method to prove that a family of functions is concentrated by showing that a subfamily is concentrated. In particular, we show that the  $i$ -th bit functions are concentrated.

## 6.1 An algorithm for finding large coefficients

We sketch how our results can be used to reduce the problem of finding the large coefficients of  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  to the case where  $n$  is an integer power of 2.

Consider the following computational problem: given a function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  and some threshold  $\tau > 0$ , recover all coefficients of  $f$  that satisfy  $|\widehat{f}(\alpha)|^2 > \tau$ . In other words, the goal is to produce a set  $\Gamma$  such that  $\alpha \in \Gamma$  if and only if  $|\widehat{f}(\alpha)|^2 > \tau$ . One way of finding these coefficients would be to calculate all the Fourier coefficients and then discard those with small magnitude. Of course, this seems wasteful if there are only a small number of large coefficients. There are algorithms which solve this problem in time  $\text{poly}(\log(n), 1/\tau, \|f\|_\infty)$  for any domain  $\mathbb{Z}_n$ , yet the case where  $n = 2^k$  has received the most attention in the literature. We refer to [1] for a recent survey on these algorithms and their practical applications. We use the name *sparse Fourier transform* (SFT) to refer to these algorithms.

We now present a new approach to this problem: instead of directly finding the large Fourier coefficients of  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  we can define  $\tilde{f} : \mathbb{Z}_{2^k} \rightarrow \mathbb{C}$  and use an existing SFT algorithm to find the large coefficients of  $\tilde{f}$ . Our results show that whenever  $f$  has large coefficients, also  $\tilde{f}$  has large coefficients. Moreover, since the proofs are constructive, if we know the set of large coefficients of  $\tilde{f}$  then this restricts where the large coefficients of  $f$  can be.

We only present an overview of the algorithm; further details can be found in the appendix. Before we describe the algorithm we mention that one can efficiently estimate the magnitude of the  $\alpha$ -th coefficient of a function  $f$  by computing

$$\left| \frac{1}{k} \sum_{i=1}^k f(x_i) \overline{\chi_\alpha(x_i)} \right|,$$

for  $x_i \in \mathbb{Z}_n$  chosen independently and uniformly at random. This is standard and follows from Chernoff's bound. The size of  $k$  depends on the required approximation, and can be fairly small for our needs.

**Main idea.** From our analysis above, we know that the large Fourier coefficients  $\tilde{f}$  are those which are close to  $\lfloor \frac{m}{n} \alpha \rfloor$  where  $|\widehat{f}(\alpha)|$  is large. Therefore, if  $|\widehat{\tilde{f}}(\beta)|$  is large we can search in a small interval around  $\lfloor \frac{n}{m} \beta \rfloor$  to find the large coefficients of  $f$ .

**Overview.** Given a function  $f : \mathbb{Z}_n \rightarrow \mathbb{C}$  and some threshold  $\tau > 0$ , we define  $\tilde{f}$  over some desired domain  $\mathbb{Z}_m$  for  $m > n$ . We say that a coefficient  $\widehat{f}(\alpha)$  is  $\tau$ -heavy if  $|\widehat{f}(\alpha)|^2 > \tau$ . If  $f$  has  $\tau$ -heavy coefficients then  $\tilde{f}$  has  $\tau'$ -heavy coefficients for some  $\tau' > 0$ . One recovers a set  $\Gamma'$  of the  $\tau'$ -heavy coefficients of  $\tilde{f}$  using an SFT.

Suppose  $\beta \in \Gamma'$ , and recall that  $\widehat{\tilde{f}}(\beta) = \sum_{\alpha \in \mathbb{Z}_n} \widehat{f}(\alpha) \langle \tilde{\chi}_{\alpha,n}, \chi_{\beta,m} \rangle$ . The set of coefficients  $\widehat{f}(\alpha)$  that significantly contribute to the magnitude of  $\widehat{\tilde{f}}(\beta)$  belongs to an interval of (small) size  $r$  around

$\widehat{f}(\lfloor \frac{n}{m} \beta \rfloor)$ . Hence, by approximating the coefficients in this interval, one can recover all the  $\tau$ -heavy coefficients (in fact, all the large coefficients) that significantly contribute to the magnitude of  $\widehat{f}(\beta)$ . We see that by translating back to  $\mathbb{Z}_n$  for each  $\tau'$ -heavy coefficient of  $\widetilde{f}$ , it is possible to find large coefficients of  $f$ .

**Issues with coefficient cancellation.** However, we need to keep in mind that some large coefficient of  $f$  may have been “canceled out” in  $\widetilde{f}$  as described in Section 5. This can be overcome, for example, by extending the search space a little bit more or by subtracting all the known large coefficients from  $f$ , and iterating.

Another way to overcome the problem of coefficients being “cancelled out” is to use the scaling property of the Fourier transform, namely if  $c \in \mathbb{Z}_n^*$  and  $h(x) = f(cx)$  then  $\widehat{h}(\alpha) = \widehat{f}(c^{-1}\alpha)$ . The scaling property of the Fourier transform can be used to first permute all the coefficients of  $f$  by defining the function  $h$  for some randomly chosen  $c \in \mathbb{Z}_n^*$ . With high probability the large coefficients of  $h$  are spread apart and so recovering them from the large coefficients of  $\widetilde{h}$  is easy using Proposition 4. We simply find  $\Gamma'$  (the set of all Fourier coefficients of  $\widetilde{h}$  that are larger than  $\tau'$ ) and isolate those  $\beta$  for which  $|\widehat{h}(\beta)|$  is the largest in some small interval (if there are any close  $\tau'$ -heavy coefficients in  $\Gamma'$ ). Then for every such  $\beta$  we test to see which one of  $|\widehat{h}(\lfloor \frac{n}{m} \beta \rfloor)|$  and  $|\widehat{h}(\lceil \frac{n}{m} \beta \rceil)|$  is large.

Once this is done, we can multiply each coefficient by  $c^{-1}$  to recover the  $\tau$ -heavy coefficients of the original function  $f$ .

## 6.2 Proving concentration

As a second application we use Theorem 1.1 to give a simple proof that a family of functions is concentrated by considering a concentrated subfamily.

**Theorem 6.1.** *Consider a family of functions  $\mathcal{F} = \{f_{2^k} : \mathbb{Z}_{2^k} \rightarrow \mathbb{C}\}_{k \in \mathbb{N}}$  and define the family  $\mathcal{F}' = \{f_n : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$ , where for each  $2^{k-1} < n \leq 2^k$  we let  $f_n(x) := f_{2^k}(x)$  for every  $x \in \mathbb{Z}_n$ . If  $\mathcal{F}$  is concentrated then  $\mathcal{F}'$  is concentrated.*

*Proof.* Consider the family  $\mathcal{G} = \{g_n : \mathbb{Z}_{m_n} \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$  where for each  $2^{k-1} < n \leq 2^k$  we let  $m_n = 2^k$  and define  $g_n := f_{2^k}$ . Since  $\mathcal{G}$  and  $\mathcal{F}$  contain the same functions,  $\mathcal{G}$  is concentrated (with the same polynomial for which  $\mathcal{F}$  is concentrated). Note that for each  $n$ , we have that  $f_n = \widetilde{g}_n$  with  $1 \leq \frac{n}{m_n} < 2$ . From Theorem 1.1, since  $\{g_n : \mathbb{Z}_{m_n} \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$  is concentrated it follows that  $\{f_n : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n \in \mathbb{N}} = \{\widetilde{g}_n : \mathbb{Z}_n \rightarrow \mathbb{C}\}_{n \in \mathbb{N}}$  is concentrated.  $\square$

**Remark 3.** *The family  $\mathcal{F}$  is taken over powers of 2 to be consistent with Theorem 1.1 above, but it could be generalised to other subfamilies.*



We now use this theorem to (re)prove that the (family of)  $i$ -th bit function is concentrated. This was first proved in [4] in a quite complicated work. Our proof is simple and follows from the following result on the subfamily of functions with domain  $\mathbb{Z}_{2^k}$ .

**Lemma 6.2.** *Let  $k \in \mathbb{N}$  and  $0 \leq i < k$ . Define  $\text{bit}_i : \mathbb{Z}_{2^k} \rightarrow \{-1, 1\}$  by  $\text{bit}_i(x) = (-1)^{x_i}$  where  $x = \sum_{j=0}^{k-1} x_j 2^j$  and  $x_j \in \{0, 1\}$ . Let  $\alpha \in \mathbb{Z}_{2^k}$ . Then  $\widehat{\text{bit}_i}(\alpha) = 0$  unless  $\alpha$  is an odd multiple of  $2^{k-i-1}$  in which case  $|\widehat{\text{bit}_i}(\alpha)| = O(2^{k-i}/|\alpha|_{2^k})$ .*

*Proof.* Writing  $N = 2^k$  we have  $\widehat{\text{bit}_i}(\alpha) = \frac{1}{N} \sum_{x=0}^{2^k-1} \text{bit}_i(x) \omega_N^{-\alpha x}$ . We write  $x = y + 2^i b + 2^{i+1} z$  where  $0 \leq y < 2^i$ ,  $0 \leq z < 2^{k-(i+1)}$  and where  $b$  is the  $i$ -th bit. Then

$$\begin{aligned} \widehat{\text{bit}_i}(\alpha) &= \frac{1}{N} \left( \sum_{y=0}^{2^i-1} \omega_N^{-\alpha y} \right) \left( \sum_{b=0}^1 (-1)^b \omega_N^{-\alpha 2^i b} \right) \left( \sum_{z=0}^{2^{k-i-1}-1} \omega_N^{-2^{i+1} \alpha z} \right) \\ &= \frac{1}{N} \left( \sum_{y=0}^{2^i-1} \omega_N^{-\alpha y} \right) (1 - \omega_N^{-2^i \alpha}) \left( \sum_{z=0}^{2^{k-i-1}-1} \omega_N^{-2^{i+1} \alpha z} \right). \end{aligned}$$

The third sum is just a sum over all  $2^{k-i-1}$ -th roots of unity, so it is  $2^{k-i-1}$  when  $\alpha$  is a multiple of  $2^{k-i-1}$  and otherwise is zero. The middle term  $(1 - \omega_N^{-2^i \alpha})$  is therefore 2 when  $\alpha$  is an odd multiple of  $2^{k-i-1}$  and is zero if it is an even multiple. For the first sum, we know from Lemma 2.3 that

$$\left| \sum_{y=0}^{2^i-1} \omega_N^{-\alpha y} \right| = \left| \frac{\sin\left(\frac{\pi}{N} \alpha (2^i - 1)\right)}{\sin\left(\frac{\pi}{N} \alpha\right)} \right| \leq \frac{1}{\left| \sin\left(\frac{\pi}{N} |\alpha|_N\right) \right|} \leq \frac{N}{\pi |\alpha|_N}. \quad (7)$$

The result then follows: the Fourier coefficient  $\widehat{f}(\alpha)$  is zero when  $\alpha$  is not an odd multiple of  $2^{k-i-1}$  and when it is non-zero it has magnitude bounded by  $2^{k-i}/|\alpha|_{2^k}$ .  $\square$

**Corollary 1.** *For every  $i \in \mathbb{N}$ , the  $i$ -th bit functions over  $\mathbb{Z}_{2^k}$ , i.e.  $\{\text{bit}_i : \mathbb{Z}_{2^k} \rightarrow \{-1, 1\}\}_{k>i}$ , are concentrated.*

**Corollary 2.** *For every  $i \in \mathbb{N}$ , the  $i$ -th bit functions  $\{\text{bit}_i : \mathbb{Z}_n \rightarrow \{-1, 1\}\}_{n>2^i}$  are concentrated.*

**Acknowledgements** We thank our supervisor Steven Galbraith for helpful assistance.

## REFERENCES

- [1] Gilbert, A.C., Indyk, P., Iwen, M., and Schmidt, L. (2014) “Recent Developments in the Sparse Fourier Transform,” in IEEE Signal Processing Magazine, **31**(5), 91–100.

- [2] Green, B. (2004) “Spectral Structure of Sets of Integers,” in Brandolini, L., Colzani, L., Travaglino, G., Iosevich, A. (eds.) Fourier analysis and convexity. Applied and Numerical Harmonic Analysis, pp. 83–96. Birkhäuser, Boston.
- [3] Laity, J. (2016) “The Sparse Fourier Transform.” MSc Thesis. The University of Auckland. Available at The University of Auckland Library, identifier 2292/29137. <http://hdl.handle.net/2292/29137>.  
Updated version at <https://www.math.auckland.ac.nz/~sgal018/JoelMSc.pdf>.
- [4] Morillo, P., and Ràfols, C. (2009) “The Security of All Bits Using List Decoding,” in Jarecki, S., Tsudik, G. (eds.) Public Key Cryptography – PKC 2009: Proceedings of the 12th International Conference on Practice and Theory in Public Key Cryptography. LNCS, vol. 5443, pp. 15–33. Springer, Heidelberg.
- [5] Shkredov, I.D. (2006) “On Sets of Large Exponential Sums,” in Doklady Mathematics, **74**(3), 860–864.
- [6] Shor, P.W. (1997) “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer,” in SIAM Journal on Computing, **26**(5), 1484–1509.

#### A. TIGHTNESS OF THE BOUND IN THEOREM 1.2

We refine Example 3 as follows. Let  $B : \mathbb{Z}_m \rightarrow \{1, -1\}$  be chosen uniformly at random and consider the function

$$f(x) := \begin{cases} B(x) & \text{when } 0 \leq x < m, \\ \chi_{\alpha, n}(x) & \text{when } m \leq x < n. \end{cases}$$

Then,

$$\widehat{f}(\alpha + j) = \frac{1}{n} \sum_{x \in \mathbb{Z}_n} f(x) \overline{\chi_{\alpha+j}(x)} = \frac{1}{n} \sum_{0 \leq x < m} B(x) \overline{\chi_{\alpha+j}(x)} + \frac{1}{n} \sum_{m \leq x < n} \omega_n^{-jx}.$$

One expects  $\sum_{0 \leq x < m} B(x)$  to be a random walk, and so with high probability  $|\sum_{0 \leq x < m} B(x)| < \sqrt{n/2}$ , hence  $|\frac{1}{n} \sum_{0 \leq x < m} B(x) \overline{\chi_{\alpha+j}(x)}| < \frac{1}{\sqrt{2n}}$ . The value  $|\frac{1}{n} \sum_{m \leq x < n} \omega_n^{jx}|$  can be shown to be large for small odd  $\pm j$ , using Lemma 2.3 (see also Eq. (7)), as well as for  $j = 0$ .

For example the value  $|\frac{1}{n} \sum_{m \leq x < n} \omega_n^{jx}|$  is 0.5 for  $j = 0$  and using Lemma 2.3, it is about  $0.636/2 = 0.318$  for  $j = \pm 1$ . Thus,  $|\widehat{f}(\alpha)| > 0.5 - \frac{1}{\sqrt{2n}}$  and  $|\widehat{f}(\alpha \pm 1)| > 0.318 - \frac{1}{\sqrt{2n}}$ . Similarly,  $|\widehat{f}(\alpha \pm 3)| > 0.106 - \frac{1}{\sqrt{2n}}$  and  $|\widehat{f}(\alpha \pm 5)| > 0.063 - \frac{1}{\sqrt{2n}}$ .

It is straightforward to show that given  $\epsilon$ , then for large enough  $n$ , we can approximate  $f$  such that  $\|f - f|_{\Gamma}\|_2^2 \leq 0.5 + \epsilon$ , for a set  $\Gamma$  of size that depends on  $\epsilon^{-1}$ . Specifically,  $f$  is  $(0.5 + \epsilon)$ -concentrated for any  $\epsilon > 0$ .

On the other hand, we have  $\tilde{f}(x) = B(x)$  and one expects this function to have no large coefficients (all coefficients bounded by  $\frac{1}{\sqrt{m}}$  as shown above). We conclude that these large coefficients cancel most of each other's magnitude in the function  $\tilde{f}$ .

From Theorem 1.2 we know that  $\tilde{f}$  is  $(2(0.5 + \epsilon) + \eta)$ -concentrated for any  $\eta > 0$ . Here the size of the set  $\Gamma'$  on which  $\tilde{f}$  is concentrated depend on  $\epsilon$  and on  $\eta$ . As we can set  $\epsilon, \eta$  to be as small as we want, we have that  $\tilde{f}$  is 1-concentrated, which trivially holds as  $\tilde{f}$  has norm 1. Moreover, as it has no large coefficients at all, and as we show in the third example under Example 1 above, with high probability it is not  $(1 - \epsilon')$ -concentrated for any  $\epsilon' > 0$ . Therefore, this example shows that the bound given in Theorem 1.2 is tight.

## B. VALUES FOR THE ALGORITHM

Let  $\Gamma \subset \mathbb{Z}_n$  be the set of all  $\tau$ -heavy coefficients of  $f$ , and suppose  $\Gamma \neq \emptyset$ . Then  $\|f - f|_{\Gamma}\|_2^2 = \|f\|_2^2 - \sum_{\alpha \in \Gamma} |\hat{f}(\alpha)|^2 \leq \|f\|_2^2 - \tau$ . By Proposition 3, there exists a set  $\Gamma' \subset \mathbb{Z}_m$  with

$$|\Gamma'| \leq |\Gamma|(|\Gamma|\|f\|_2^2/\epsilon' + 3) \quad (8)$$

such that

$$\|\tilde{f} - \tilde{f}|_{\Gamma'}\|_2^2 \leq \frac{n}{m}\|f\|_2^2 - \tau + \epsilon' + 2\sqrt{(1 - \tau)\epsilon'}. \quad (9)$$

We choose an appropriate  $\epsilon'$  such that the right-hand side is non-trivial (for example, if  $\tau < \frac{n}{m}\|f\|_2^2/2$  we can make the right-hand side equal  $\frac{n}{m}\|f\|_2^2 - 2\tau$ ). We calculate a suitable threshold  $\tau'$  for the function  $\tilde{f}$ . Since  $\|f\|_2^2 = \sum_{\alpha \in \mathbb{Z}_n} |\hat{f}(\alpha)|^2$ , then  $f$  can have at most  $\|f\|_2^2/\tau$  coefficients larger than  $\tau$ . In other words,

$$|\Gamma| \leq \|f\|_2^2/\tau. \quad (10)$$

From (9) we have

$$\|\tilde{f}|_{\Gamma'}\|_2^2 = \|\tilde{f}\|_2^2 - \|\tilde{f} - \tilde{f}|_{\Gamma'}\|_2^2 \geq \frac{n}{m}\|f\|_2^2 - \frac{n}{m}\|f\|_2^2 + \tau - \epsilon' - 2\sqrt{(1 - \tau)\epsilon'} = \tau - \epsilon' - 2\sqrt{(1 - \tau)\epsilon'}.$$

It follows that in  $\Gamma'$  there is at least one coefficient  $\hat{\tilde{f}}(\beta)$  such that

$$|\hat{\tilde{f}}(\beta)|^2 \geq \frac{\tau - \epsilon' - 2\sqrt{(1 - \tau)\epsilon'}}{|\Gamma'|} \geq \frac{\tau - \epsilon' - 2\sqrt{(1 - \tau)\epsilon'}}{(\|f\|_2^2/\tau) ((\|f\|_2^2/\tau)\|f\|_2^2/\epsilon' + 3)},$$

using (8) and (10). We set  $\tau'$  to be the latter value.